Aalto University
School of Science

CS-E4640: Advanced Topics in Software Systems

Machine Learning Project Management

Phuong Pham
Researcher of Aaltosea
phuong.pham@aalto.fi

ML Project Management

Phuong Pham
Researcher at Aaltosea
phuong.pham@aalto.fi

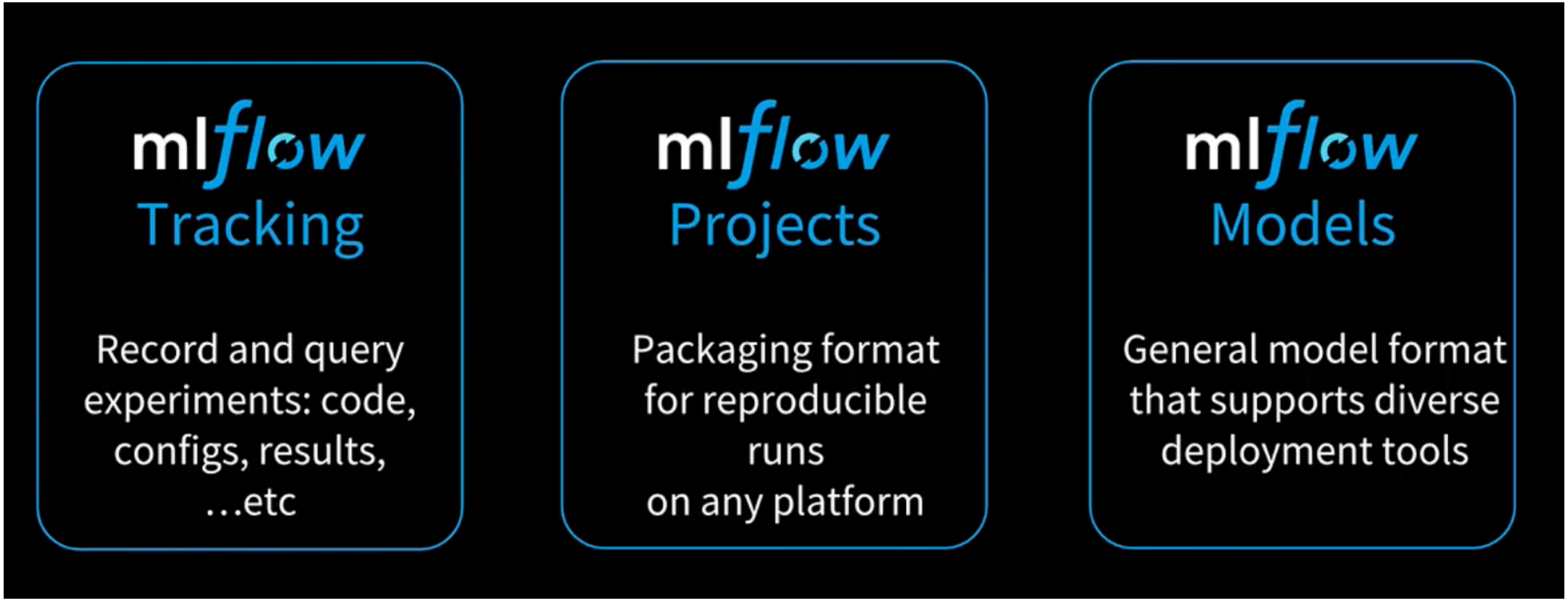
Motivation

How to collect the experimental data?

Machine Learning Applications and Systems

How to manage a machine learning experiment?

MLFlow



mlflow
Tracking

Record and query experiments: code, configs, results, …etc

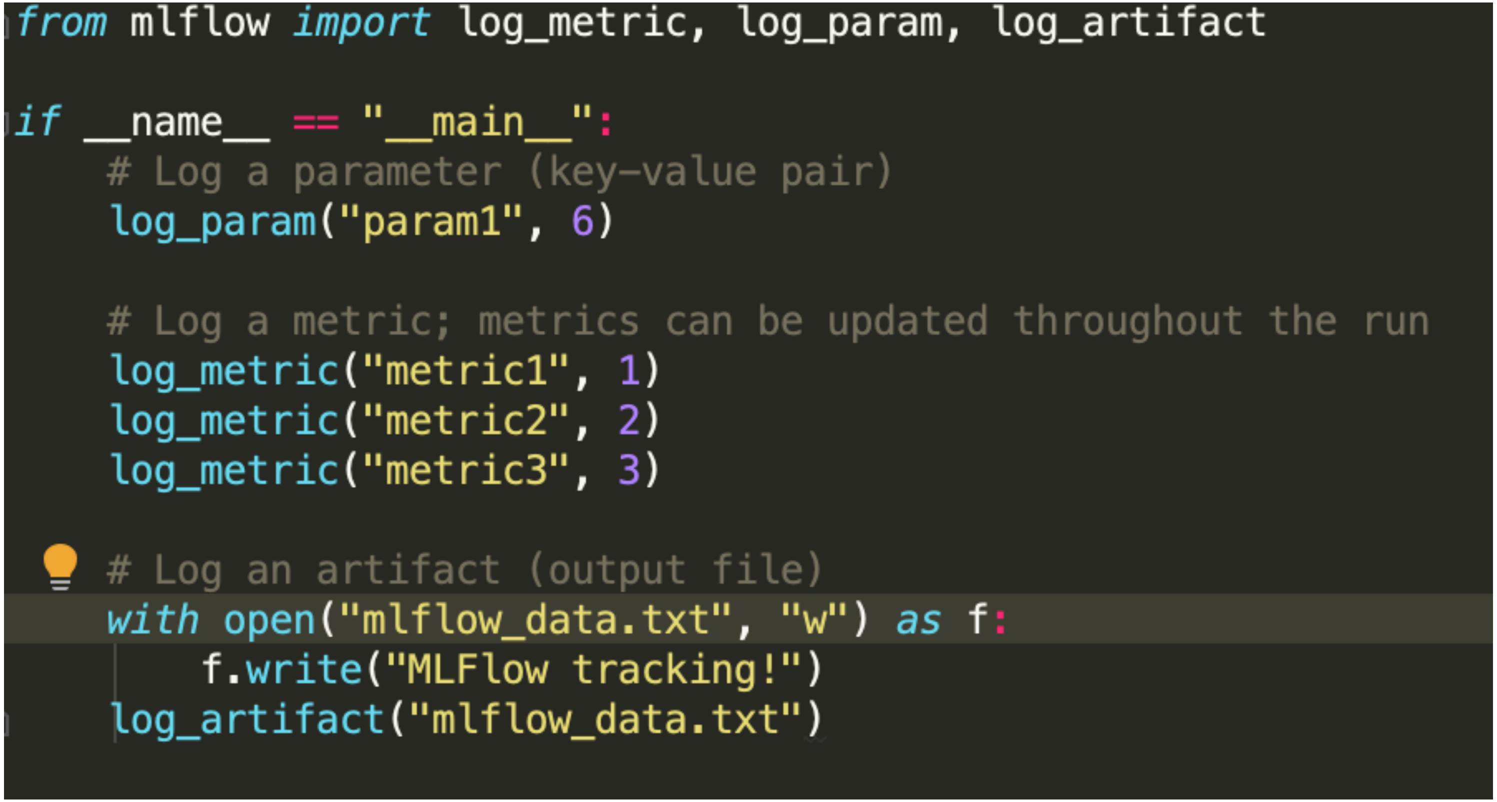
mlflow
Projects

Packaging format for reproducible runs
on any platform

mlflow
Models

General model format that supports diverse deployment tools

Tracking



```python
from mlflow import log_metric, log_param, log_artifact

if __name__ == "__main__":
    # Log a parameter (key-value pair)
    log_param("param1", 6)

    # Log a metric; metrics can be updated throughout the run
    log_metric("metric1", 1)
    log_metric("metric2", 2)
    log_metric("metric3", 3)

    # Log an artifact (output file)
    with open("mlflow_data.txt", "w") as f:
        f.write("MLFlow tracking!")
    log_artifact("mlflow_data.txt")
```

Projects

Diverse set of tools

- Keras, Pytorch, R, Tensorflow, Spark, …

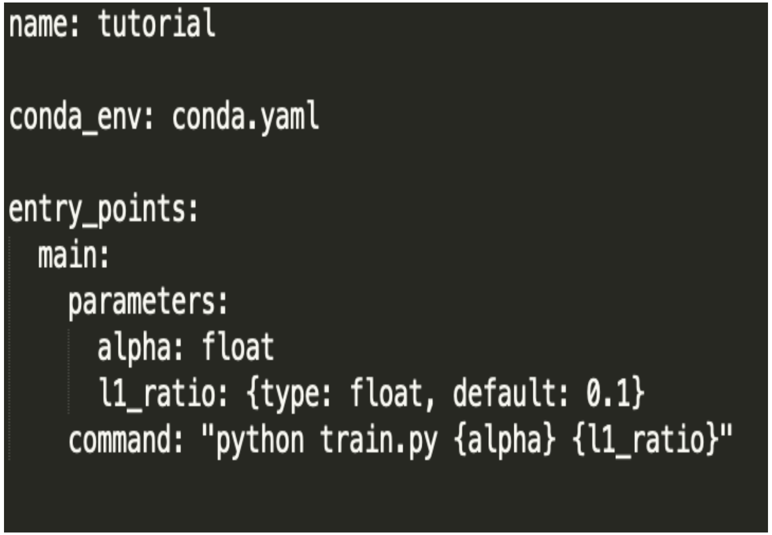
Diverse set of environments

- Linux, Mac, Window, …

→ It is difficult to productionalize

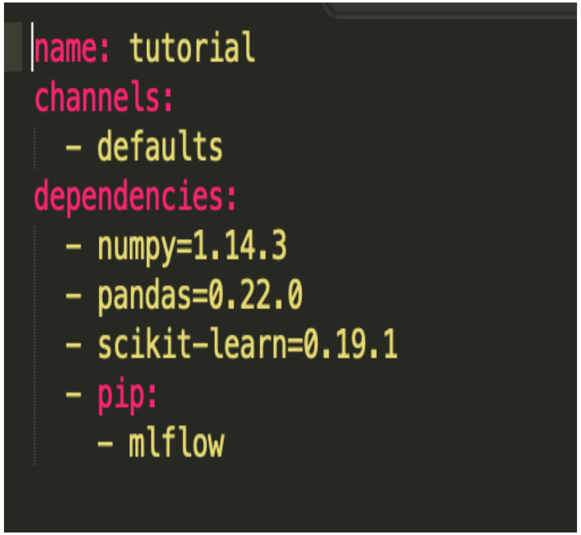
Projects

MLProject file



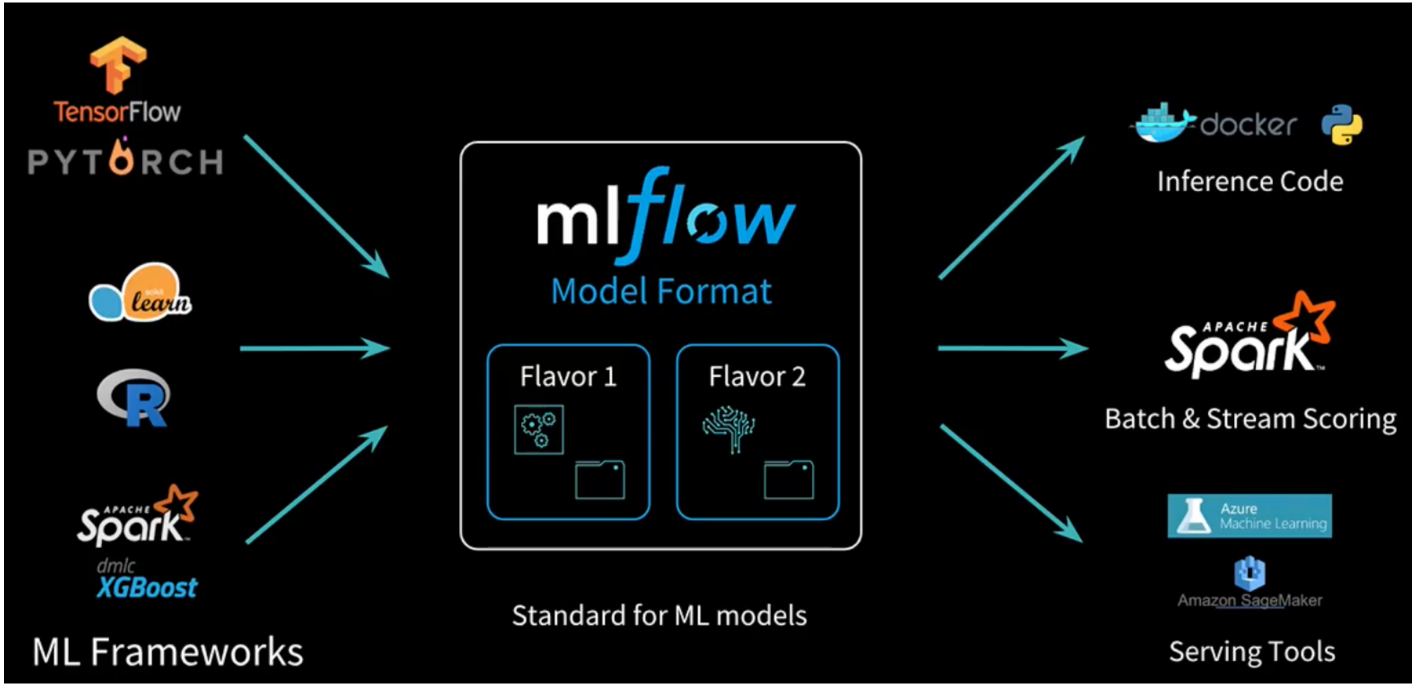
```
name: tutorial

conda_env: conda.yaml

entry_points:
  main:
    parameters:
      alpha: float
      l1_ratio: {type: float, default: 0.1}
    command: "python train.py {alpha} {l1_ratio}"
```

conda.yaml



```
name: tutorial
channels:
  - defaults
dependencies:
  - numpy=1.14.3
  - pandas=0.22.0
  - scikit-learn=0.19.1
  - pip:
    - mlflow
```

Models



A MLflow Model is a standard format for packaging machine learning models

Models

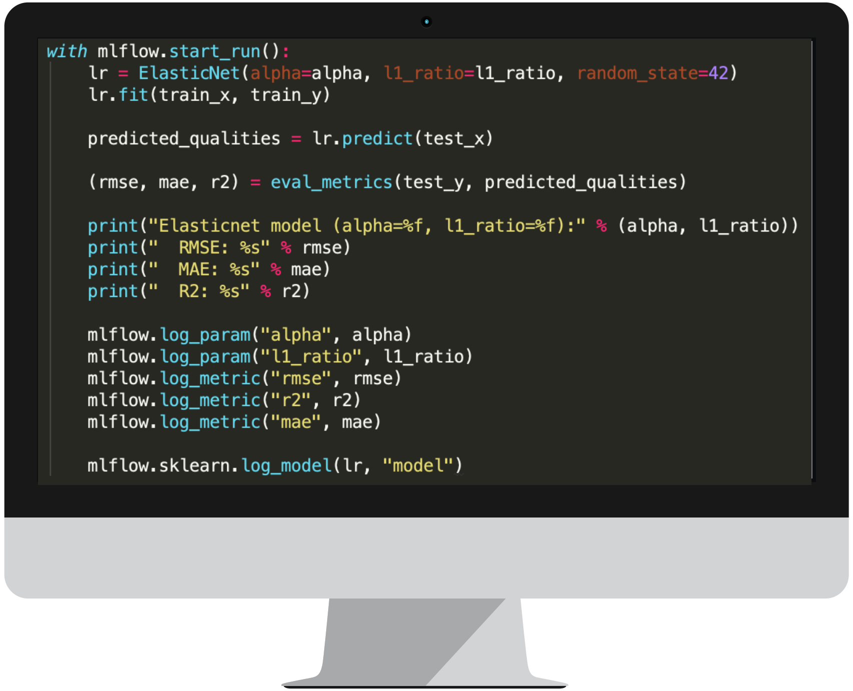
- Serving the model with mlflow:

```
$ mlflow models serve –m /path-to-the-mlflow-model/ -p app_deployment_port
```

- Access the model for predicting test data:

```
$ curl –X POST –H “Content of data” http://server:app_deployment_port/invocations
```

Summary



```python
with mlflow.start_run():
    lr = ElasticNet(alpha=alpha, l1_ratio=l1_ratio, random_state=42)
    lr.fit(train_x, train_y)

    predicted_qualities = lr.predict(test_x)

    (rmse, mae, r2) = eval_metrics(test_y, predicted_qualities)

    print("Elasticnet model (alpha=%f, l1_ratio=%f):" % (alpha, l1_ratio))
    print("  RMSE: %s" % rmse)
    print("  MAE: %s" % mae)
    print("  R2: %s" % r2)

    mlflow.log_param("alpha", alpha)
    mlflow.log_param("l1_ratio", l1_ratio)
    mlflow.log_metric("rmse", rmse)
    mlflow.log_metric("r2", r2)
    mlflow.log_metric("mae", mae)

    mlflow.sklearn.log_model(lr, "model")
```

Your ML Projects

Enjoy managing your mlprojects with mlflow

Discussion